Journal of Translational Medicine

RESEARCH

Open Access

stRGAT: identifying spatial domains in spatial transcriptomics via a relational graph attention network

Zhengqian Zhang[1], Jialiang Wang[1], Junjun Ren[1], Yiduo Zhao[1], Xu Yang[2], Jingyi Bai[3], Shuang Liu[1] and YongZhuang Liu[1*]

Abstract

Background Spatial transcriptomics has rapidly advanced the study of tissue architecture by integrating gene expression profiles, spatial coordinates, and high-resolution histology images. However, accurately identifying spatial domains with coherent transcriptomic and morphological patterns remains challenging.

Methods We developed stRGAT, a deep learning framework that integrates spatial information and gene expression to learn low-dimensional latent embeddings for spatial domain identification. stRGAT adaptively models the similarity among neighboring and cross-regional spots, enabling coherent spatial partitioning. Additionally, the framework can be extended to detect spatially variable genes that reflect biologically meaningful spatial expression patterns.

Results We conducted extensive evaluations on cross-species datasets, continuous tissue sections, and fine-grained mouse brainstem datasets. Quantitative comparisons against current mainstream methods demonstrate that stRGAT more accurately identifies spatial domains, achieving higher clustering agreement metrics and clearer domain boundaries. Furthermore, stRGAT's ability to capture spatially variable genes underscores its versatility in spatial transcriptomic analysis.

Conclusions stRGAT provides a robust and generalizable approach for spatial domain delineation and associated tasks in spatial transcriptomics, offering improved performance over existing methods and broad applicability for studying tissue organization and function.

Keywords Identify spatial domains, Spatial transcriptomics, Graph neural network, Spatially variable gene detection, Relational graph attention

*Correspondence:
YongZhuang Liu
yongzhuang.liu@hit.edu.cn
[1]School of Computer Science and Technology, Harbin Institute of Technology, 92 West Dazhi Street, Harbin, Heilongjinang 150001, China
[2]School of Electronics and Information Engineering, Harbin Institute of Technology, 92 West Dazhi Street, Harbin, Heilongjiang 150001, China
[3]School of Mathematics, Harbin Institute of Technology, 92 West Dazhi Street, Harbin, Heilongjiang 150001, China

Introduction

Many different types of cells are present among the specific tissues of multicellular creatures; each cell has specific functions, and the cycles and behaviors of the cells are related to their surrounding environments. In particular, determining the location of cells and spatial information about tissues can aid in effectively researching the biological functions and diseases of tissues. Single-cell RNA sequencing (scRNA-seq) has become one of the core technologies in the field of genomics because of its impressively high resolution, and it plays an important role in the analysis of cell groups in immunology, biomedicine and other disciplines [1]. However, scRNA-seq involves mixing tissue cells during the sequencing process; this process leads to a lack of cell location information, which is vital for understanding tissue functions. To address this issue, the spatial differential transcript (SRT) sequence was selected by the Nature journal as the 2020 method of the year [2]. The development and research concerning spatial transcription groups, such as 10x Visium [3], Slideseq [4] and other technologies, provides important guarantees for understanding the spatial dimensions and tissue functions of cells.

Identifying the spatial domain (that is, the region where the gene expression location and histological location are the same) is one of the most basic and important research topics in spatial transcriptomics [5]. However, identifying accurate and biologically functional spatial domains is still a great challenge. The methods for identifying spatial domains can be divided into spatial clustering and nonspatial clustering methods. Nonspatial clustering methods mainly use traditional clustering methods, such as the k-means and Louvain algorithms [6]; however, these methods consider only sparse gene expression data, and it is difficult for the clustering results to correspond with tissue sections. Spatial clustering methods yield strongly interpretable analysis results by combining gene expression data, spatial location information and morphological images and fully considering spatial dimension information.

For example, the main inspiration of BayesSpace [7] came from image processing technology, in which a Markov random field (MRF) was used to implement a Bayesian model to encourage spots in the same cluster to be close to one another. StLearn [8] extracts features from tissue images, and it standardizes gene expression data through the distance between adjacent points in space before performing clustering. The updated data are subsequently utilized to identify groups that exemplify the transcriptional profiles of certain cell types and phenotypes. The StLearn approach involves using the pseudospace-time (PST) distance measure to calculate transcriptional states and reconstruct spatial transition gradients. This is accomplished by considering the connections within and between cell types, as well as clusters, through a directed minimum spanning tree optimization technique. The SpaGCN [9] uses a graph convolutional network (GCN) to process gene expression, spatial location, and morphology data. By generating an undirected weighted graph, the gene expression information of adjacent points is aggregated, and an unsupervised clustering algorithm is used for clustering, thereby identifying spatial domains. SEDR [10] uses a deep autoencoder to construct a latent gene representation in a low-dimensional latent space and subsequently embeds the corresponding spatial information through a variational graph autoencoder. RESEPT [11] maintains 3D embeddings based on graph neural networks [12] and maps them to the RGB image color channel. Then, using image segmentation, a supervised convolutional neural network is used to segment the RGB image. The purpose of recognizing the spatial domain is thus achieved. Although these methods can consider the spatial characteristics of spatial transcriptomics, node similarity is calculated before training and cannot be adaptively learned; moreover, these methods cannot fully explore the spatial structural characteristics of spatial transcriptomics data. The scalability of the algorithms is poor, and different datasets lead to worse results. In addition, most algorithms focus on identifying accurate spatial domains, but research on the fine identification of specific tissues is lacking [13]. Many methods consider the similarity between adjacent nodes. For a neural network model, too much consideration of the similarity between adjacent nodes may cause the model to ignore the relationships at different levels. More importantly, for deep learningbased methods, extensions to multiple consecutive sections yield greater algorithmic optimization and biological value.

For these reasons, we developed stRGAT to accurately identify spatial domains. The process involves incorporating spatial information and gene expression profiles to acquire low-dimensional latent embeddings. stRGAT integrates gene expression, spatial location, and histological image data to accurately identify spatial domains in spatial transcriptomics adaptively via relational graph neural networks. We conducted extensive testing on the 10x Visium dataset and compared it with the existing algorithms. This dataset includes commonly used human dorsolateral prefrontal cortex (DLPFC)data within the field, cross-species data, and continuous slice data, proving its advantage in identifying downstream analysis tasks in the spatial domain.

Methods

Data description

stRGAT is applied to 10x Visium-generated data. The first dataset consists of 12 consecutive brain stem slice

samples acquired from three adult human donors, specifically, the six-layered human DLPFC [14]. The white matter (WM) and DLPFC layers were manually annotated by the authors. Each point corresponds to an average of 3,462 unique molecular indices (UMIs) and an average of 1,734 genes.

The second dataset is from the Mouse Kidney Section (Coronal) dataset, which has a tissue slice resolution of 10μm and contains 1,438 spots. The median number of genes per spot is 6,516. We use the corresponding anatomical Allen Mouse Brain Atlas [15] to more accurately compare the results of the experiment.

When analyzing continuous slice tasks, we use Mouse Brain Serial Section "Introduction" (Sagittal-Anterior) and Mouse Brain Serial Section "Introduction" (Sagittal-Posterior). Mouse Brain Section (Sagittal-Anterior) contains 2,695 spots, and the median number of each spot is 6,228. Mouse Brain Section (Sagittal-Posterior) contains 3,355 spots, and the average number of each spot is 4,772.

Data processing

The input data for stRGAT consist of gene expression data and histological images. Points outside the main tissue regions must be removed. Then, common methods are used for the logarithmic transformation and standardization of the gene expression data. We refer to the processing strategy of the SpaGCN algorithm.

The spatial gene expression data are held in an $N \times D$ matrix, where N represents the number of spots and D represents the number of genes. The coordinates of each spot in two-dimensional space (x,y) are also included. Genes that are expressed in fewer than three spots are excluded.

To normalize the gene expression values for each spot, the UMI count for each gene is divided by the total UMI count across all genes in that spot, multiplied by 10,000, and then converted to the natural logarithm scale. We use gene expression data and histological images as input data. These data are read and converted into matrix formats and subsequently regulated.

Converting SRT data into graph-structured data

After performing data preprocessing, the gene expression and histology image data must be transformed into graph-structured data. The input data are first transformed into an undirected band-weighted graph.

In the spatial transcriptome data, for two fixed points, the physical locations and histological information of the two nodes are equally important. First, the physical position similarity of the tissue section is calculated. Histological images have regionspecific characteristics. To improve the identification effect attained for the boundary area, we use the Gaussian smoothing method to calculate the color value of the current node to avoid the effect of a single outlier.

$$g\left(x_p, y_p\right) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

In the formula, (x_p, y_p) is the coordinate of a pixel, and σ is the standard deviation. For an undirected graph $G\left(V, E\right)$, we can calculate the similarity between two points. Since the distributions of organizational histological images, gene expression matrices, and positional coordinates are different, we use the standardized Euclidean distance measure to calculate the distance between different nodes V.

$$x_i' = \frac{x_i - u_i}{s_i} \tag{2}$$

u_i is the average of all the data in this dimension, and s_i is the corresponding standard deviation. The specific operation calculates the standardized Euclidean distance between points in stRGAT. Each graph vertex represents a point, and an edge with a specified weight is used to connect each pair of points. stRGAT reduces the dimensionality of the preprocessed gene expression matrix according to principal component analysis (PCA). The top 50 principal components are used as inputs, and this strategy works well for all the datasets analyzed in this paper. For two nodes u and v, the formula for calculating the distance between them is as follows.

$$d\left(u, v\right) = \sqrt{\sum_{k=1}^{n} \left(\frac{x_i - y_i}{S_i^2}\right)^2} \tag{3}$$

Relational graph construction

For each relation $r \in \{\text{spa,exp,his}\}$, we define edge weights based on relation-specific distances:

$$w_{ij}^{(r)} = \exp\left(-\frac{d_{ij}^{(r)}}{\sigma_r}\right). \tag{4}$$

Distances follow

$$\begin{aligned} d_{ij}^{(\text{spa})} &= ||\mathbf{s}_i - s_j||_2, \\ d_{ij}^{(\text{exp})} &= ||\mathbf{x}_i - \mathbf{x}_j||_2, \\ d_{ij}^{(\text{his})} &= 1 - \cos\left(\mathbf{h}_i, \mathbf{h}_j\right). \end{aligned} \tag{5}$$

Each relation is row-normalized as $\tilde{A}^{(r)} = D^{(r)-1} A^{(r)}$, with $D_{ii}^{(r)} = \sum_j A_{ij}^{(r)}$. A fused adjacency for clustering is

$$A^{\text{fuse}} = \sum_r \lambda_r \tilde{A}^{(r)}, \quad \lambda_{\text{spa}} = 1, \lambda_{\text{exp}} = \lambda_{\text{his}} = 0.2\,(\text{default})\,. \tag{6}$$

Histology feature extraction

Co-registered H&E patches (40×) are Macenko-normalized and encoded by a ResNet18 pretrained on ImageNet. The 512-d raw embedding is linearly projected to 64-d for scale alignment:

$$\mathbf{h}_i = \mathbf{W}_h \mathbf{h}_i^{\text{raw}} \in \mathbb{R}^{64}. \tag{7}$$

The projected $\mathbf{h}_i$ is used for histology relations and can be concatenated with expression features during reconstruction.

Relational graph attention autoencoder

The relational graph attention network takes normalized gene expression data as its input, with the input set defined as $\mathbf{h} = \left\{\vec{h}_1, \vec{h}_{i+1}, \ldots, \vec{h}_N\right\}$, where $i \in \{1, 2, 3, 4, \ldots, N\}$ and N represents the number of nodes. The output nodes are $\mathbf{h}' = \left\{\vec{h}'_1,\ \vec{h}'_{i+1}, ..., \vec{h}'_N\right\}$, with $i \in \{1, 2, 3, 4, \ldots, N\}$. We use Relational Graph Attention (ARGAT) [16].

For each relation *r*, node *i* and its neighborhood are combined with their intermediate representations to form the logit $E_{i,j}^{(r)}$. The attention coefficients e_{ij} are generated from the logit matrix using the softmax function. These attention coefficients are used to calculate a weighted sum of the nodes in the neighborhood for each relation. The process involves gathering these elements together and then applying a nonlinearity mechanism to generate the revised representation of node *i*.

For relation *r*, attention coefficients incorporate learned logits and normalized adjacency:

$$\alpha_{ij}^{(r)} = \frac{\exp\left(\text{LeakyReLU}\left(a_r^{\mathrm{T}}\left[\mathbf{W}_r\mathbf{h}_i^{(l)} \middle\| \mathbf{W}_r\mathbf{h}_i^{(l)}\right]\right)\right)}{\sum_{k\in N_r(i)}\exp\left(\text{LeakyReLU}\left(a_r^{\mathrm{T}}\left[\mathbf{W}_r\mathbf{h}_i^{(l)} \middle\| \mathbf{W}_r\mathbf{h}_i^{(l)}\right]\right)\right)}\tilde{A}_{ij}^{(r)} \tag{8}$$

Multi-head, multi-relation aggregation is

$$\mathbf{h}_i^{l+1} = \Big\|_{m=1}^{M}\ \sigma\left(\sum_r \sum_{j\in N_r(i)} \alpha_{ij}^{(r,m)} \mathbf{W}_r^{(m)} \mathbf{h}_j^{(l)}\right) \tag{9}$$

where σ is ELU/ReLU and $\|$ denotes concatenation. The total loss remains $\mathrm{L} = \lambda_{\text{feat}}\mathrm{L}_{\text{feat}} + \lambda_{\text{adj}}\mathrm{L}_{\text{adj}} + \lambda_{\text{kl}}\mathrm{L}_{\text{kl}}$ as in the main text.

To make the network sufficiently capable of converting the input features into more complex features, we need to have at least one adaptable linear transformation. First, we use a common linear transformation defined by a weight matrix, represented as $W \in \mathbb{R}^{F'\times F}$, where F is the number of features in each node.

Self-attention [17] is applied to each node individually, followed by the implementation of a shared attention mechanism on the nodes. The attention coefficient calculation method is as follows:

$$\mathrm{e}_{ij} = a\left(\mathbf{W}\vec{h}_i,\ \mathbf{W}\vec{h}_j\right) \tag{10}$$

To ensure comparability, we normalize the similarity weights using a softmax function via the following approach:

$$g_{ij} = \text{softmax}_j\ (e_{ij}) = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(e_{ik}\right)} \tag{11}$$

To adaptively learn the similarity between adjacent points, we adopt a self-attention mechanism that is widely used in graph neural networks. The attention mechanism g is a single-layer feedforward neural network whose parameters are shared among the nodes and weighted by a weight vector. The unfolded attention mechanism is represented as follows:

$$g_{ij} = \frac{\exp\left(\text{Re LU}\left(\vec{a}^T\left[\mathbf{W}\vec{h}_i \middle\| \mathbf{W}\vec{h}_j\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{Re LU}\left(\vec{a}^T\left[\mathbf{W}\vec{h}_j \middle\| \mathbf{W}\vec{h}_k\right]\right)\right)} \tag{12}$$

T represents transposition, and $\|$ is the concatenation operation. The normalized attention coefficients are utilized to calculate a linear combination of the corresponding features. These combined features are subsequently chosen as the final output features for each node.

$$\vec{t}_i NL\left(\sum_{j\in\mathcal{N}_i} g_{ij}\mathbf{w}\vec{h}_j\right) \tag{13}$$

NL represents nonlinearity.

Clustering

Based on the output of the graph attention network, stRGAT uses an unsupervised clustering algorithm [18] for clustering, which identifies different points in the spatial domain. The specific employed clustering method is Louvain.

Spatially variable genes (SVGs) identification

We detect spatially variable genes on Visium slides with co-registered H&E images. Spots with fewer than 500 detected genes or mitochondrial fraction above 20% are removed, followed by library-size normalization and log1p transform. H&E tiles are Macenko-normalized; a ResNet18 encoder (ImageNet pretrain, 40× patches) is linearly projected to 64 d to match expression feature scale.

Highly variable genes are selected with variance stabilisation; PCA keeps 50–100 PCs depending on dataset size. Candidate genes must appear in at least 5% of retained spots and exceed the 60th percentile of dispersion. We build a multimodal graph with three edge sets: (i) spatial kNN on Euclidean spot coordinates, (ii) histology affinity from cosine similarity of ResNet embeddings, (iii) expression mutual kNN in PC space. For spatial statistics we collapse these relations into a symmetric weight matrix $W = [w_{ij}]$ via row-normalised summation to satisfy Moran's I assumptions.

For gene g with centered expression vector $\mathbf{x}$, Moran's I is

$$I_g \frac{n}{W_+} \frac{\sum_{i \neq j} w_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \quad W_+ = \sum_{i \neq j} w_{ij} \tag{14}$$

which captures global spatial autocorrelation. We also compute a Gini-based spatial variance

$$G_g = \frac{\sum_{i \neq j} w_{ij} |x_i - x_j|}{2W + \mu_x}, \ \mu_x = \frac{1}{n} \sum_i x \tag{15}$$

highlighting local high-expression patches (GiniClust-inspired). Both scores are retained to capture complementary spatial signals.

For each gene we permute spatial labels 1,000 times (graph fixed) to form empirical nulls for I_g and G_g. Two-sided p-values p_I, p_G are combined by Fisher's method,

$$\chi^2_{\text{Fisher}} = -2 (\log p_I + \log p_G) \tag{16}$$

and adjusted by Benjamini–Hochberg to control FDR at 5%. Genes passing FDR are ranked by a composite z-score $s_g = \alpha z (I_g) + (1 - \alpha) z (G_g)$ ($\alpha \in [0,1]$, default 0.5), balancing global and local spatial structure.

Hyperparameter tuning of stRGAT

The hyperparameters used in stRGAT may vary in different experiments. Taking the DLPFC dataset as an example, the hyperparameters include the use of the adaptive moment estimation optimizer [19] with an initial learning rate of 1×10^{-4} and a weight decay set to 1×10^{-3}. The default number of iterations is 250. The hardware setup consists of an NVIDIA GeForce RTX 3090 graphics card, an Intel(R) Core(TM) i9-12900KS CPU, and 64 GB of memory.

Benchmarking

The ARI is used to evaluate and compare the effectiveness of various clustering methods on datasets that have spot-type labels. The key characteristics of the ARI include the following:

$$\text{ARI} = \frac{\sum_{ij} \binom{N_{ij}}{2} - \frac{\left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2}\right]}{\binom{N}{2}}}{\frac{1}{2}\left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2}\right] - \frac{\left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2}\right]}{\binom{N}{2}}} \tag{17}$$

Result

Overview of stRGAT. stRGAT first processes the input SRT data, as shown in Fig. 1. stRGAT uses the input data to construct graph-structured data; that is, graph vertices are constructed for spatial positional information and organizational images, and attention weight relationships are established between the points. Then, stRGAT uses a relational graph attention network to adaptively learn the gene expression information of adjacent points. Unlike traditional graph convolutional neural networks, stRGAT uses attention mechanisms, which can learn the similarity between adjacent points. Finally, stRGAT uses unsupervised clustering methods to cluster the expression matrix output by the relational graph attention network to achieve spatial domain recognition. To fully demonstrate the advantages of stRGAT, we conducted extensive comparative tests on multiple datasets and utilized the current mainstream algorithms.

To quantitatively evaluate the spatial recognition capability of the stRGAT, we first tested it on the 10x Visium dataset containing 12 human DLPFC images. Utilizing the adjusted Rand indices [1] as clustering accuracy metrics, we compared our method with five current mainstream spatial methods (SpaGCN, stLearn, BayesSpace, SEDR and STAGATE). Since most of the previously developed algorithms have proven that spatial domain recognition algorithms are generally superior to nonspatial methods [20], we only compared our approach with more mature spatial methods. With respect to a publicly available dataset called the 10x Visium ST benchmarking dataset [14], annotated the cortical layers (L1–L6) and WM of 12 DLPFC slides using gene markers and cytoarchitecture, manually marking them as shown in Fig 2a.

With respect to the human DLPFC dataset, we first compared the performance of the tested algorithm on

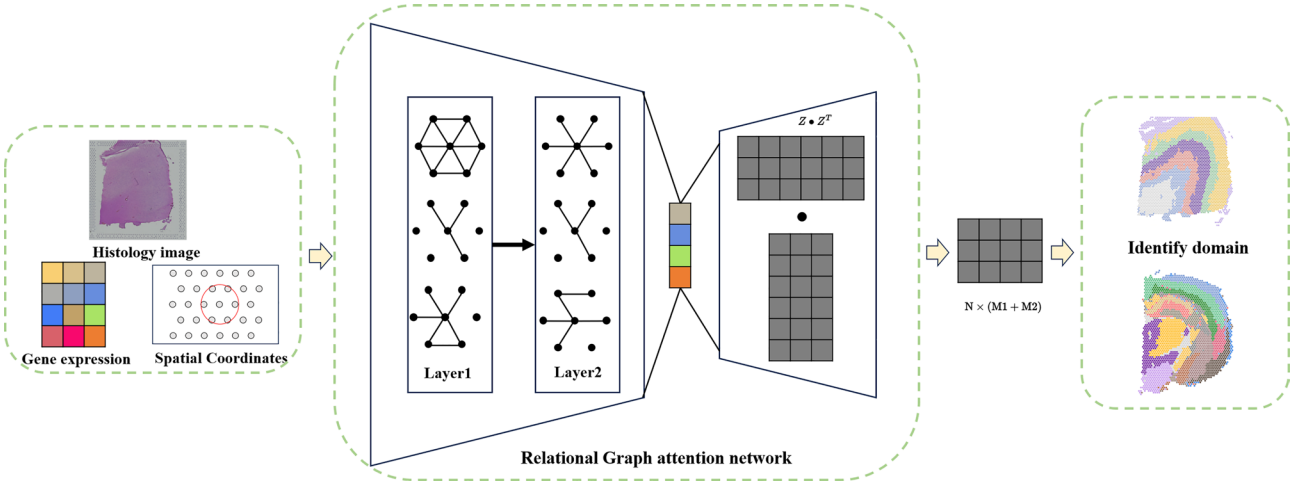

Fig. 1 Workflow of the stRGAT algorithm. The workflow of stRGAT starts with an ST data examination, prioritizing gene expression data, histological images, and spatial location information as inputs. Next, the data are dimensionally reduced and normalized, and an undirected graph is constructed using the standardized Euclidean distance to measure the similarity between nodes. The preprocessed data are subsequently fed into a relational graph attention network to extract the final latent embedding, and unsupervised clustering algorithms are used for spatial domain identification

the 151673 dataset. To ensure the fairness of the experiment, we did not choose to use the default parameters when running the other algorithm models because default parameters could not ensure that the other algorithms would achieve optimal experimental results. We made appropriate adjustments to ensure that the comparative algorithms could reach or approach the optimal results displayed in their papers. Figure 2c shows the spatial domain recognition results produced by BayesSpace, SEDR, stLearn, SpaGCN, STAGATE and stRGAT on dataset 151673. Compared to the other methods, stRGAT could more effectively identify the expected cortical layer structure. Among the five models, the spatial domain identified by stRGAT clearly depicted the layer boundaries and achieved the highest clustering accuracy (ARI = 0.537). Its overall results were also closer to the manual annotation results. In contrast, although the recognition results of BayesSpace were relatively uniform, they were also affected by the utilized algorithm, which led to unclear cortical layering and discontinuity of the identified spatial domain, resulting in a larger gap between the clustering results and manual annotations. Relatively speaking, SEDR had a better layer closure effect, but this effect was not good for recognizing individual cortical layers, and more outliers were present. StLearn produced obvious cortical layer clustering errors, and there were more outliers that differed greatly from the actual cortical layering results. SpaGCN benefited from the structure of the utilized graph convolutional neural networks, and its experimental results were ideal; however, regions with unreasonable layering remained. Moreover, SpaGCN even identified one less spatial domain result. STAGATE also used a diagram of an attention network to model the information of adjacent nodes. Although it obtained a high ARI score (0.57), it did not use histological image information, and it also considered adjacent node information and ignored a long-distance node. This information caused a layering error.

To further validate the performance of stRGAT, we conducted thorough tests on 12 datasets of human DLPFC data, recorded the experimental results obtained on each dataset, and calculated the average ARI values. The experimental results are shown in Fig. 2b. The box plot shows that stRGAT produced a higher ARI, and its ARI distribution was more uniform, which means that the performance of stRGAT was more stable. The average ARIs achieved by BayesSpace, SEDR, stLearn, SpaGCN, STAGATE and stRGAT on the 12 datasets were 0.378, 0.410, 0.369, 0.390, 0.409, and 0.419, respectively. This experiment proves that stRGAT is superior to the existing state-of-the-art methods. The average ARI of stRGAT was 0.419, which was higher than that of the best-performing SEDR approach (0.410).

To verify the ability of stRGAT to handle spatial domain recognition task, we tested it using a mouse hippocampal dataset. Before testing, we first labeled the hippocampal structure as shown in Fig. 3. (the leftmost H&E staining image). To better represent the hierarchical layered structure of the hippocampus, we used the Allen Brain Reference Atlas images annotated by The Allen Institute for Brain Science, as shown in the middle image of Fig. 3. The shape of the hippocampus, as well as its hierarchical layered structure, can be clearly observed. The rightmost image in Fig. 3 shows the spatial domain results recognized by stRGAT. We can see clearly that stRGAT accurately identified the main structures of the hippocampus, and the recognized tissue types were relatively

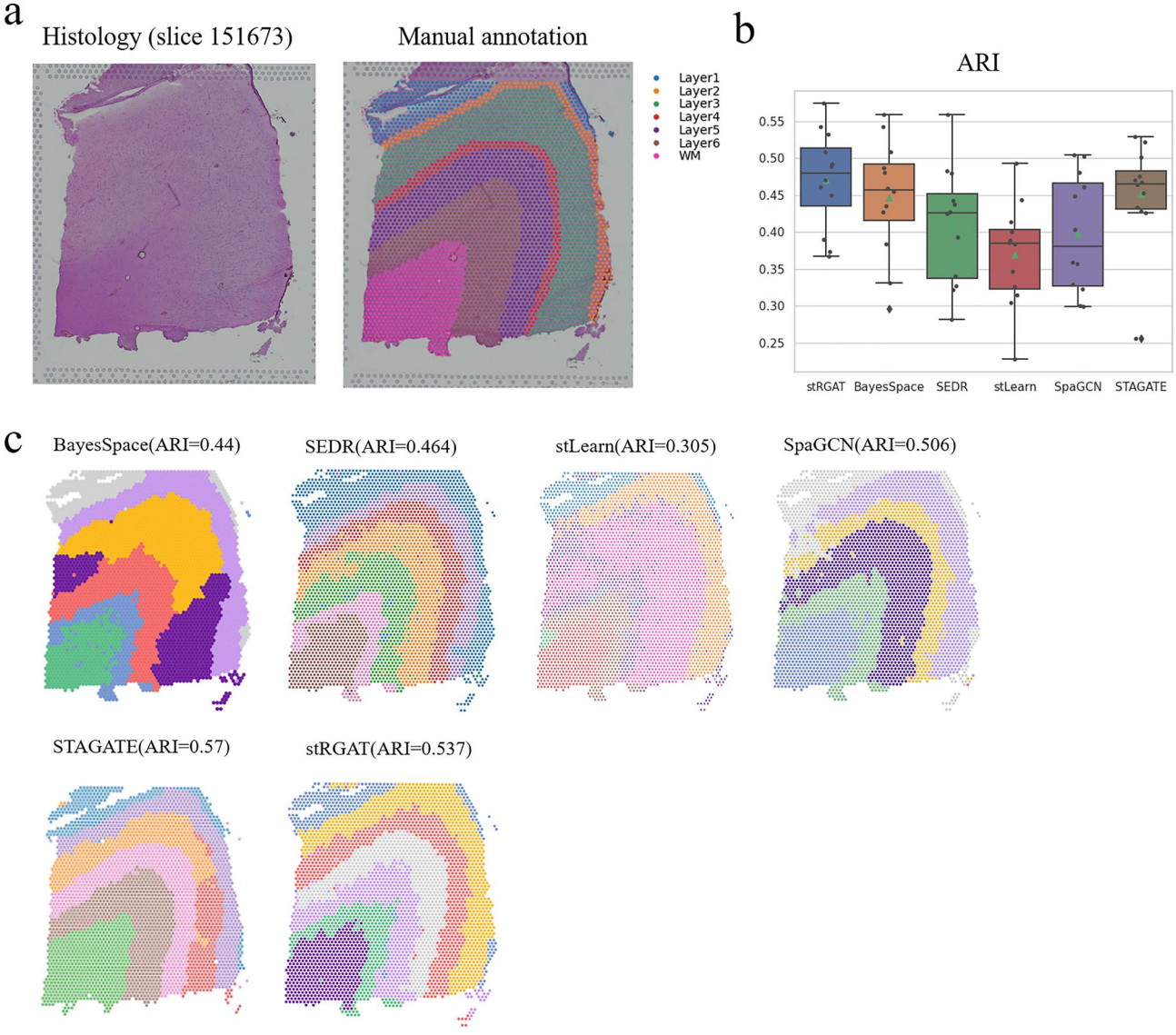

Fig. 2 stRGAT achieves an enhanced ability to recognize layer formations that are present in human DLPFC tissue. **a** Ground-truth spots of cortical layers and WM in the DLPFC (section 151673). **b** The ARI boxplots produced for the 12 datasets contained in the DLPFC. The median and upper and lower quartiles are shown by the centerline, box limits and whiskers in a boxplot. The whiskers represent 1.5 times the interquartile range. **c** Based on the experimental results obtained on dataset 151676, the compared algorithms are SpaGCN, stLearn, BayesSpace, STAGATE and SEDR

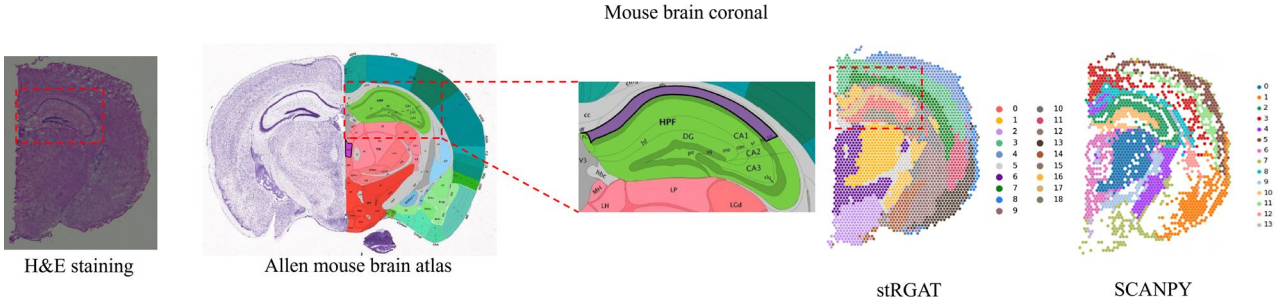

Fig. 3 Spatial domains of the coronal regions of mouse brain tissue. H&E staining was performed on the raw data (left). The corresponding anatomical Allen mouse Brain Atlas (middle, https://atlas.brain-map.Org/) was utilized. The spatial domains identified by stRgaT and Scanpy are also shown (right). The red box denotes the cornu ammonis and dentate gyrus areas in the coronal portion

continuous, with basically no abnormal points produced; this is consistent with the tissue hierarchical structure annotated by the Allen Brain Reference Atlases. Furthermore, in terms of detail processing, stRGAT identified four tissue layers, namely, 'Dentate gyrus, molecular layer', 'Field CA1, stratum lacunosum-meleculare', 'Field CA1, pyramidal layer' and 'Field CA1, stratum oriens', demonstrating the excellent performance of stRGAT. We compared stRGAT with the SCANPY [21] algorithm, as shown in the figure. It is evident that stRGAT more completely recognized the hippocampus than SCANPY, and in terms of detailed processing, stRGAT identified narrower tissue structures. The recognition effect of SCANPY was unsatisfactory because the recognized tissue of the hippocampus was not continuous and obviously differed from the tissue types annotated by the Allen Brain Reference Atlases.

The above experimental analysis focused only on a single tissue slice. We further explored the ability of stRGAT to detect mutations in multiple consecutive tissue slices using mouse anterior section and mouse posterior section datasets. These two datasets were derived from a mouse brain, with the anterior and posterior regions adjacent to each other. Figure 4 shows that for consecutive tissue slices in the same region, stRGAT could accurately infer the clustering relationship between the two tissues. In other words, in adjacent regions of two tissues, stRGAT could accurately infer that the same tissue belongs to the same category. The figure shows that stRGAT identified continuous spatial domains across tissues, which is highly important for further exploring the spatial attributes of spatial transcriptomics. Moreover, as the amount of data increased, the spatial domain identification effect was also enhanced. In the experiments conducted on the 151,672 dataset, we first validated the ability of stRGAT to identify spatial domains. Compared with several representative methods, including SpaGCN, STAGATE, SEDR, BayesSpace, PROST, and stLearn, our model achieved the highest Adjusted Rand Index (ARI) of 0.60, surpassing the previous best result of 0.59 reported by PROST [22]. This demonstrates the strong performance of stRGAT in spatial domain identification (See Fig. 5).

To evaluate the individual and combined contributions of each modality (gene expression, spatial coordinates, and histology), we performed ablation experiments on the stRGAT framework. Specifically, we constructed models using only gene expression features, gene expression combined with spatial information, and the full set of modalities including histology features. For each configuration, we assessed the quality of the resulting spatial domain clustering by comparing the inferred clusters against known anatomical layers using clustering agreement metrics such as Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). This analysis enables a quantitative understanding of how each modality contributes to domain discrimination and how its integration improves the overall performance. The specific results are shown in Fig. 6. (1) Joint expression+spatial achieves the best agreement with ground truth (ARI = 0.525, NMI = 0.690). (2) Expression-only degrades performance (ARI = 0.352, NMI = 0.496), showing spatial cues are necessary. (3) Spatial-only is weakest (ARI = 0.149, NMI = 0.385), indicating expression is the dominant signal, but spatial context substantially boosts layer delineation when combined. Overall, the gain from expression+spatial over either single modality demonstrates complementary contributions.

To further evaluate the capability of stRGAT in detecting spatially variable genes (SVGs), we identified 4560 SVGs using our model. A comparison with the top 50 SVGs selected by the PROST method revealed a high degree of consistency in the computed PI scores. These results confirm the effectiveness of stRGAT in accurately identifying biologically relevant SVGs.

The boxplots in Fig. 7 compare the distribution of Moran's I statistics across three methods—stRGAT, SpatialDE [23], and SpaGCN—on the DLPFC 151672 dataset. Among these methods, stRGAT exhibits the highest median and overall distribution of Moran's I values, indicating the strongest spatial autocorrelation captured by the model. SpatialDE demonstrates a moderate level of

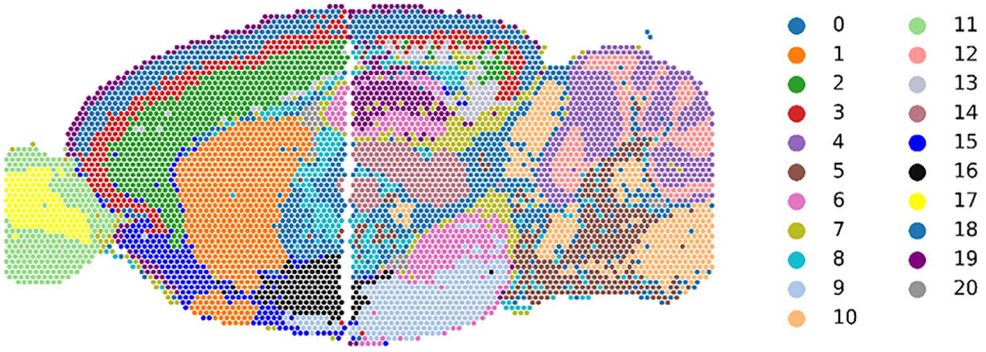

Fig. 4 Joint spatial domain detection results obtained across multiple mouse brain tissue sections using stRGAT

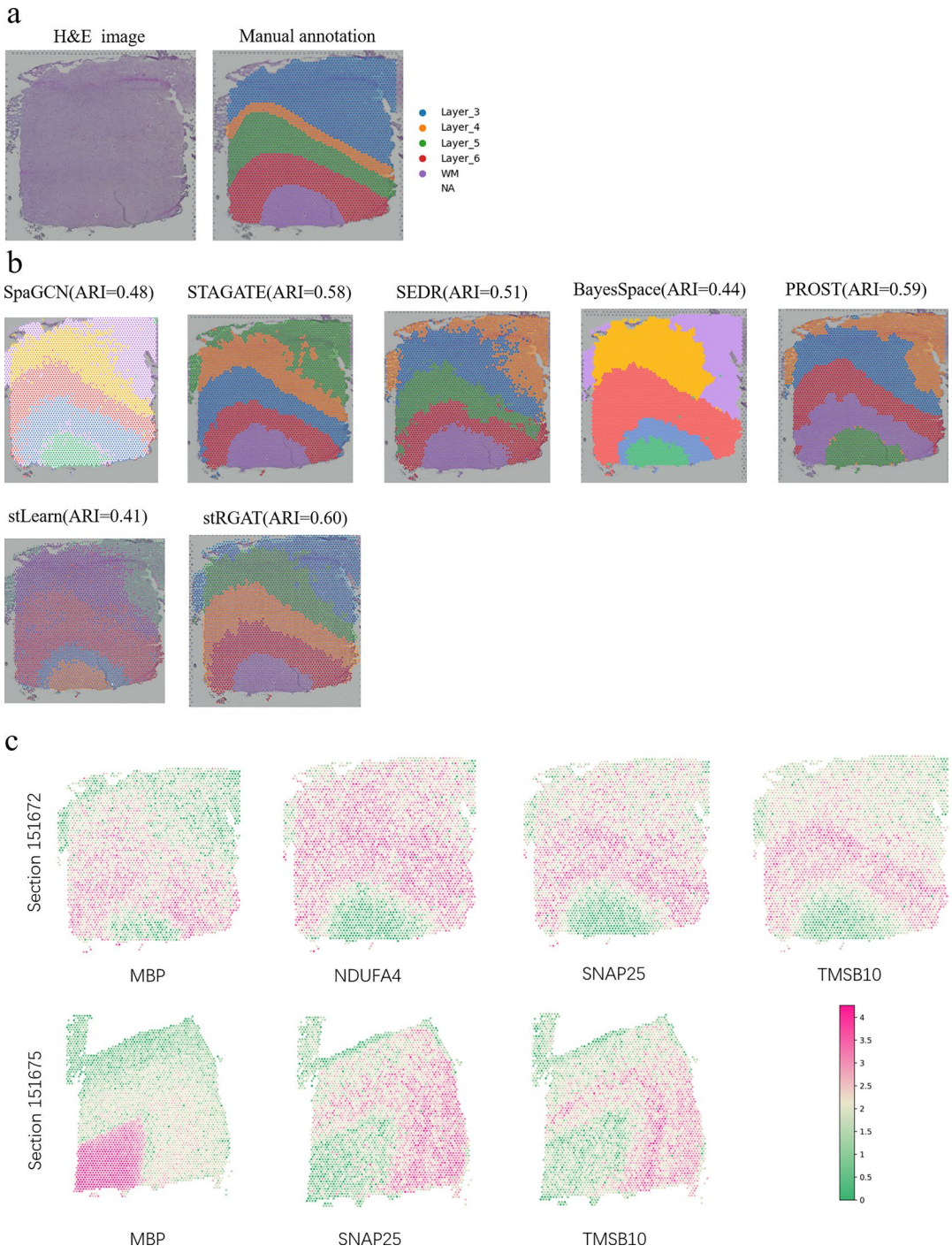

Fig. 5 stRGAT reaches sota level in spatial domain recognition task on 151672 dataset. **a** Groundtruth spots of cortical layers and WM in the DLPFC (section 151672). **b** Based on the experimental results obtained on dataset 151672, stRGAT . **c** Show the same spatial patterns in the DLPFC section 161672 and151675, demonstrating the transportability of SVGs

spatial structure, while SpaGCN yields somewhat lower Moran's I values. These results suggest that stRGAT more effectively captures intrinsic spatial patterns in the data, whereas the alternative methods retain only moderate spatial dependence.

Discussion

In the field of biological analysis based on spatial transcriptomics data, accurately identifying spatial domains is crucial for understanding tissue organization and functional specialization of cell populations. In this study, we proposed stRGAT, a relational graph attention–based

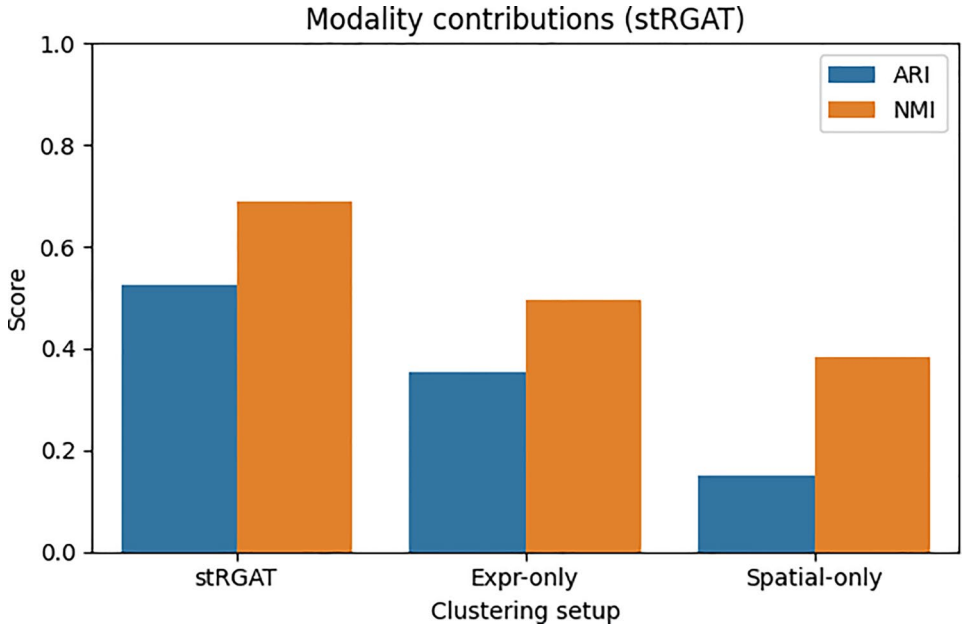

Fig. 6 stRGAT modality ablation on DLPFC 151673. Bars show ARI and NMI for expression+spatial (stRGAT), expression-only, and spatial-only. Joint modeling of expression and spatial coordinates yields the highest agreement with annotated layers, evidencing that spatial signals enhance expression-based clustering

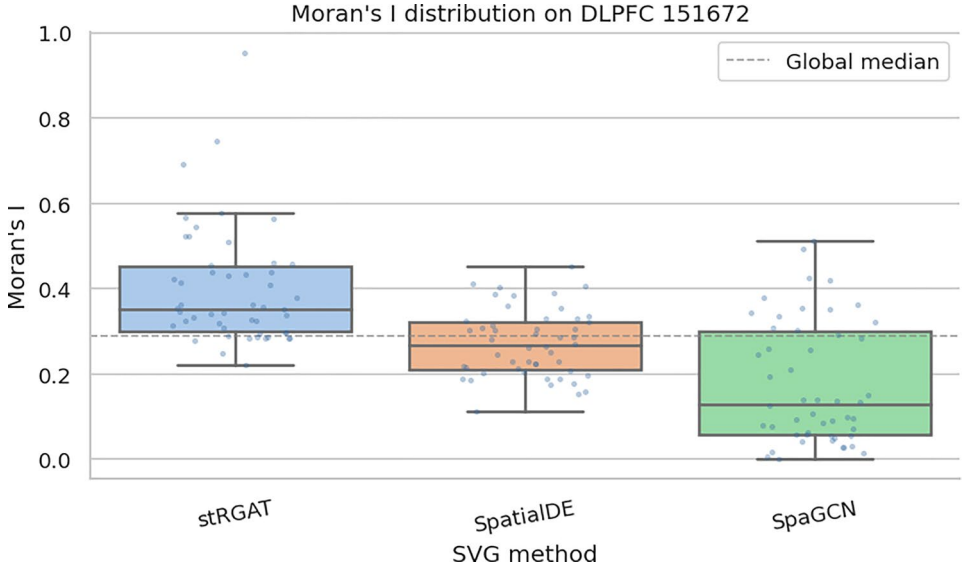

Fig. 7 Moran's I distribution of three SVG methods on DLPFC 151672

framework that jointly leverages gene expression profiles, spatial coordinates and histology images to learn low-dimensional embeddings for spatial domain detection. Across multiple 10× Visium datasets, including human DLPFC sections and mouse brain coronal and sagittal slices, stRGAT consistently recovered anatomically meaningful layers and fine-grained tissue architectures, such as hippocampal subfields, with higher ARI and clearer domain boundaries than several mainstream methods. Furthermore, by jointly modeling anterior and posterior mouse brain sections, stRGAT demonstrated its ability to handle continuous tissue sections, and increasing the amount of spatial transcriptomics data further improved the robustness and resolution of inferred domains.

Spatial transcriptomics is rapidly evolving as a cornerstone technology for biology and medicine, and recent reviews have highlighted the growing role of deep learning methods in coping with the complexity of spatially resolved data [24] and the expanding landscape

of analytical strategies for spatial transcriptomics data analysis [25, 26]. In addition, comprehensive surveys of spatial transcriptomic technologies have documented breakthroughs in measurement technologies and their integration with other omics and imaging modalities [27, 28]. Embedding stRGAT within such integrative frameworks may enable refined delineation of rare or transitional cell states and improve the interpretation of domain boundaries in terms of underlying cell-type compositions. Integration of spatial transcriptomic data and single-cell transcriptomic data has been shown to generate high-resolution maps of cellular subpopulations and contextualize tissue microenvironments [29–31], further underscoring opportunities to combine spatial domain detection with transcriptomic resolution enhancement. In parallel, advancements in deep learning–based image analysis and segmentation offer promising avenues to enhance histology-guided domain delineation and multimodal representation learning [10, 11].

Despite its promising performance, stRGAT has several limitations suggesting directions for future work. First, our evaluation focuses on a limited set of 10× Visium datasets, and we have not yet systematically assessed the generalizability of stRGAT to other spatial transcriptomics technologies or large-scale cohorts. Second, the current graph construction strategy relies on predefined spatial and expression-based neighborhoods; although the relational attention mechanism can mitigate some suboptimal edges, GAT-based models are generally more sensitive to noisy inputs and spurious connections than simpler graph neural network architectures. Third, our assessment mainly relies on clustering metrics and qualitative inspection of domain boundaries, and does not fully exploit recent advances in explainability and attribution for graph neural networks [32]. A broader benchmarking of stRGAT against additional spatial domain detection and spatially variable gene identification methods, including interpretable multi-modal frameworks [33] and convolutional network–based learning strategies [34], would provide a more comprehensive understanding of its relative strengths and weaknesses.

These limitations naturally suggest several directions for future development. From a methodological perspective, we plan to enhance the robustness of stRGAT by incorporating denoising strategies, adversarial training or regularization terms that explicitly penalize unstable attention patterns on noisy or low-coverage spots. In addition, developing modules that expose node- and edge-level relational attributions, in the spirit of recent interpretable multi-modal frameworks for spatial transcriptomics [33], could help elucidate which genes, relations and histological features drive specific spatial partitions. Another promising direction is to explore learnable or data-driven graph construction schemes that more comprehensively integrate spatial proximity, histological context and expression similarity.

In summary, stRGAT provides a flexible and powerful relational graph attention framework for spatial domain identification that can adaptively learn multi-relational spatial representations from gene expression, spatial coordinates and histology images. With the rapid accumulation of high-resolution and multi-modal spatial transcriptomics datasets, we expect that stRGAT and related graph-based approaches will play an increasingly important role in uncovering spatial cellular organization, guiding downstream analyses such as spatially variable gene detection, trajectory inference and cell–cell interaction analysis, and contributing to a deeper understanding of tissue architecture and disease mechanisms.

Acknowledgements
Not applicable.

Author contributions
Z.Z. and Y.L. conceived the study and supervised the overall project. Z.Z. designed the stRGAT model, implemented the full pipeline, conducted all computational experiments, and performed the result analysis. J.W. and J.R. contributed to model optimization and robustness testing across datasets. Y.Z. and X.Y. conducted data preprocessing, graph construction, and contributed to the implementation of evaluation metrics. J.B. assisted with benchmarking against existing spatial domain detection methods and contributed to figure preparation (Figs. 2–5). S.L. carried out the literature review and contributed to editing the manuscript draft. Z.Z. and Y.L. wrote the main manuscript text. All authors read, discussed, and approved the final manuscript.

Funding
This work was supported by Science and Technology Innovation 2030 -"Cancer, Cardiovascular and Cerebrovascular, Respiratory and Metabolic Diseases Prevention and Treatment Research" Major Project (Project No.2023ZD0506403) by Medical and Health Science and Technology Development Research Center, National Health Commission of People's Republic of China, National Natural Science Foundation of China (No:62072140) and Heilongjiang Provincial Science and Technology Department.
2022ZX103C01).

Data availability
All the datasets used in this paper are published datasets that are available for download. (1) Human DLPFCs within the spatialLIBD (http://spatial.libd.org/spatialLIBD/); (2) Mouse brain section-coronal regions (https://www0.10xgenomics.com/resources/datasets/mouse-kidney-section-coronal-1-standard-1-1-0); (3) Mouse brain section- sagittalanterior and sagittal-posterior regions (https://www0.10xgenomics.com/resources/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-1-0). (4) The stRGAT algorithm is implemented in Python and is available on Github (https://github.com/leader402/stRGAT).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 September 2025 / Accepted: 30 December 2025

Published online: 14 January 2026

References

1. Suvà ML, Tirosh I. Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. Mol Cell. 2019;75(1):7–12.
2. Marx V. Method of the Year, 2020: spatially resolved transcriptomics. Nat Methods. 2021;18(1):9–14.
3. Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. Cell. 2020;182(2):497–514.
4. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363(6434):1463–67.
5. Zhang L, Chen D, Song D, Liu X, Zhang Y, Xu X, et al. Clinical and translational values of spatial transcriptomics. Signal Transduct Target Ther. 2022;7(1):111.
6. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008;2008(10):P10008.
7. Zhao E, Stone MR, Ren X, Pulliam T, Nghiem P, Bielas JH, et al. BayesSpace enables the robust characterization of spatial gene expression architecture in tissue sections at increased resolution. bioRxiv. 2020. https://doi.org/10.1101/2020.09.13.295790.
8. Pham D, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cellcell interactions and spatial trajectories within undissociated tissues. bioRxiv. 2020. https://doi.org/10.1101/2020.05.31.125658.
9. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021;18(11):1342–51.
10. Xu H, Fu H, Long Y, Ang KS, Sethi R, Chong K, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med. 2024;16(1):12.
11. Chang Y, He F, Wang J, Chen S, Li J, Liu J, et al. Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. Comput Struct Biotechnol J. 2022;20:4600–17.
12. Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. In: Proceedings of the IEEE International Joint Conference on Neural Networks. vol. 2; 2005. p. 729–34.
13. Zahedi R, Ghamsari R, Argha A, Macphillamy C, Beheshti A, Alizadehsani R, et al. Deep learning in spatially resolved transcriptomics: a comprehensive technical view. Brief Bioinform. 2024;25(2):bbae082.
14. Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci. 2021;24(3):425–36.
15. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res. 2012;41(D1):D996–1008.
16. Busbridge D, Sherburn D, Cavallo P, Hammerla NY. Relational graph attention networks. arXiv. 2019, 1904.05811.
17. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv. 2014, 1409. 0473.
18. Xie J, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning; 2016. p. 478–87.
19. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv. 2014;1412.6980.
20. Xu C, Jin X, Wei S, Wang P, Luo M, Xu Z, et al. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Res. 2022;50(22):e131.
21. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. https://doi.org/10.1186/s13059-017-1382-0.
22. Liang Y, Shi G, Cai R, Yuan Y, Xie Z, Yu L, et al. PROST: quantitative identification of spatially variable genes and domain detection in spatial transcriptomics. Nat Commun. 2024;15(1):600.
23. Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. Nat Methods. 2018;15:343–46.
24. Heydari AA, Sindi SS. Deep learning in spatial transcriptomics: learning from the next next-generation sequencing. Brief Bioinform. 2023;25(2):bbae082.
25. Du J, et al. Advances in spatial transcriptomics and related data analysis strategies. J Transl Med. 2023;21:350.
26. Li R, et al. Navigating the landscapes of spatial transcriptomics data analysis: current approaches and future perspectives. Brief Bioinform. 2024;25:bbac097.
27. Chen T, You L, Hardillo JU, Chien M. Spatial transcriptomic technologies enable measurement of gene expression throughout tissue space and its applications. Front Biol. 2023;18:1–23.
28. Hu W, Zhang Y, Mei J, et al. Spatial transcriptomics in human biomedical research and clinical application. Curr Med. 2023;2(6):1–15. https://doi.org/10.1007/s44194-023-00023-4.
29. Garcia-Alonso L, Lorenzi V, Mazzeo CI, Alves-Lopes JP, Roberts K, SanchoSerra C, et al. Single-cell roadmap of human gonadal development. Nature. 2022;607(7919):540–47.
30. Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, et al. Spatial multi-omic map of human myocardial infarction. Nature. 2022;608(7924):766–77.
31. Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat Rev Genet. 2021;22(10):627–44.
32. Yuan H, Yu H, Gui S, Ji S. Explainability in graph neural networks: a taxonomic survey. IEEE Trans Pattern Anal Mach Intell. 2022;45(5):5782–99.
33. Zong Y, Yu T, Wang X, Wang Y, Hu Z, Li Y. conST: an interpretable multi-modal contrastive learning framework for spatial transcriptomics. bioRxiv. 2022. https://doi.org/10.1101/2022.01.13.476201.
34. Xu Y, McCord RP. CoSTA: unsupervised convolutional neural network learning for spatial transcriptomics analysis. BMC Bioinf. 2021;22(1):397.

Publisher's Note